

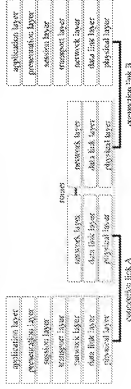
Chapter 5: Network Layer

1. Basics
2. Interconnection networks
3. Switching Technologies
4. Routing
5. Example network protocols: IP, X.25

1

1 Basics (cont'd)

- A **router** operates on the network layer:
 - it can act as a junction between two or more networks to realize data transmissions among them
 - it allows the coupling of networks with different transmission media and different data link protocols



1 Basics

- The network layer provides the means to transport messages from a source node over an interconnection network (IN) with several intermediate nodes to a destination node
- major functions in this layer:
 - routing, i.e. the selection of a transport path either *static* (fixed for all packets of a message) or *dynamic* (new path for each packet)
 - detection of overload situations
 - support of a logical addressing scheme
- famous example: IP

2 Interconnection networks

- each IN can be considered as a graph (V, E) :
 - node set V consists of input/output nodes and internal switches
 - edge set $E \subseteq V \times V$ contains all interconnection links
 - each edge is characterized by a certain data width d and a bit rate $r \Rightarrow$ the resulting data rate is $b = d \cdot r$
 - the degree of a node is the number of incoming and outgoing edges
 - a path defines a way through a graph from an input node i to an output node j
 - the maximum distance between any two nodes in the graph is called *diameter* (measured in the number of *hops*)
 - the *cost* c is typically defined as the number of edges in the graph
- the graph determines the *topology* of a network

2 Interconnection networks (cont'd)

- some criteria for the evaluation of INs:
 - transfer rate r_{single} of a single connection link (often also called "link bandwidth")
 - network throughput r_{total} in a network of c connection links:

$$r_{\text{total}} = c \cdot r_{\text{single}}$$
 (often also called "network bandwidth")
 - average or maximum latency between two network nodes
 - the bisection width w is the minimum number of connections links that must be cut through to divide the network into two subnetworks of approx. equal size
 - the bisection throughput $r_{\text{bisection}}$ is the throughput through the bisection cut: $r_{\text{bisection}} = w \cdot r_{\text{single}}$ (often also called "bisection bandwidth")

2 Interconnection networks (cont'd)

- one simple approach consists in coupling N DTEs by a network with star topology:
 - all DTEs are coupled by point-to-point links to a central node
 - the central node can be
 - 1) a computer (DTE): all connections are realized in the DTE by hard- and software, central node can become a bottleneck
 - ⇒ star network is a static network
 - 2) a hub: each incoming message is broadcasted to all nodes
 - ⇒ star network is a broadcast network
 - 3) a switch: each incoming message is sent only to the destination node, many messages as possible should be switched simultaneously
 - ⇒ star network is a dynamic network

2 Interconnection networks (cont'd)

- an interconnection network is built for connecting a certain number of DTEs (*Data Terminal Equipments*, e.g. telephones or computers)
- each IN can be realized as
 - a static network: the DTEs are directly coupled by point-to-point interconnection links
 - a broadcast network: all DTEs are connected to a shared transmission medium
 - a dynamic or switched network: the DTEs are connected via switches that allow to realize reconfigurable paths through the network

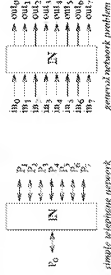
2.1 Static interconnection networks

- some topologies for static networks:
 - full mesh: ideal network, each of the N nodes is directly connected to all other nodes, requires $N(N-1)/2$ links ⇒ impracticable
 - partial mesh
 - star
 - tree
 - fat tree
 - grid
 - linear array
 - ring
 - hypercube
- each network of N nodes with static topology can be described by a (in general binary) $N \times N$ interconnection matrix



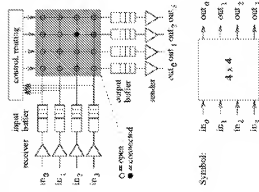
2.2 Dynamic interconnection networks

- Each dynamic interconnection network consists of internal communication links and switches
- there is at least one path from each input i to each output j
- each bidirectional dynamic interconnection network with N ports can be considered as an unidirectional dynamic network with N inputs and N outputs:



2.2 Dynamic interconnection networks (cont'd)

- an $N \times N$ crossbar is a single-stage non-blocking dynamic IN
- architecture, here for $N=4$:
- internal data width w
- trivial routing algorithm (the destination in the packet header is the index of the crossbar output)
- allows the realization of arbitrary permutations (in principle also all broadcast connection patterns are possible)
- high cost: $O(N^2)$

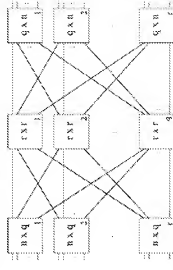


2.2 Dynamic interconnection networks

- the cost of a dynamic network is often defined as the number of binary switches (*crosspoints*)
- dynamic networks can be classified as
 - non-blocking: each connection from a free input i to a free output j can always be realized
 - blocking: a path from a free input i to a free output j can not be realized, if a required internal connection link is already busy
 - rearrangeable: each connection from a free input i to a free output j can be realized by rearranging one or several other already existing paths

2.2 Dynamic interconnection networks (cont'd)

- Idea: The number of switches can be reduced by constructing multi-stage networks of small crossbar switches
- Architecture of a three-stage Clos network: (Charles Clos, 1953)



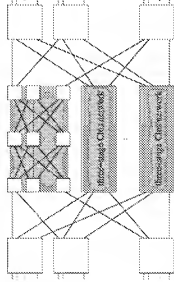
- A Clos network is non-blocking, if $q \geq 2n-1$

2.2 Dynamic interconnection networks (cont'd)

- Cost of a three-stage Clos network with N inputs/outputs:

$$C_{\text{Clos3}} = 6N^{3/2} \sim 3N \rightarrow O(N^{3/2})$$

- Construction of a five-stage Clos network:

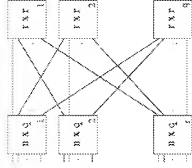


- Cost of a five-stage Clos network:

$$C_{\text{Clos5}} = 16N^{4/3} \sim 14N + 3N^{2/3} \rightarrow O(N^{2/3})$$

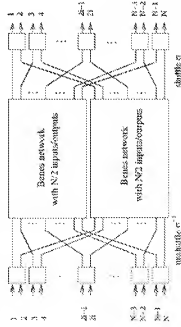
2.2 Dynamic interconnection networks (cont'd)

- Clos networks represent one of the most important dynamic network architectures
- In practice, Clos networks are realized with bidirectional interconnection links
⇒ one-sided Clos network
- same characteristics as a two-sided unidirectional Clos network (because it can be regarded as two-sided Clos network folded at the middle vertical axis)
⇒ also non-blocking for $q \geq 2n - 1$



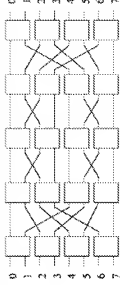
2.2 Dynamic interconnection networks (cont'd)

- A Clos network with $n = q$ is called a Beneš network (named after V. Beneš, 1962)
- It is typically shown in its recursive representation with binary switches ($n = 2$ and $N = 2^q$);



2.2 Dynamic interconnection networks (cont'd)

- Beneš network for $N = 8, n = 2$



- Cost: $C_{\text{Beneš}} = 2N \cdot (2 \log_2 N - 1)$

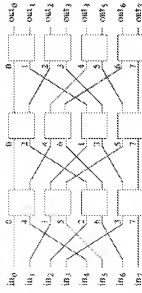
- A Beneš network is rearrangeable:

- smallest network that allows the realization of all $N!$ permutations of the N inputs onto the N outputs
- however a complex graph-theoretic algorithm required for the calculation of the switch positions

2.2 Dynamic interconnection networks (cont'd)

• Omega network (Lawrie, 1975):

- log N switch stages
- in each stage there are $m = N/k$ crossbars of size $k \times k$
- the interconnection between two switch stage is either a butterfly (see next slide) or a shuffle function
- Example:** Omega network for $n = 8, k = 2$, with shuffle σ



2.2 Dynamic interconnection networks (cont'd)

- simple routing algorithm there for $k=2$: bit d_i of the binary destination $d = (d_{n-1} \dots d_1 d_0)$ defines, which output must be used along the path in network stage i
- it is a blocking network, but many permutations can be realized
- Example:** Omega network for $n = 8, k = 2$, with butterfly interconnect
- Omega network can also be constructed for arbitrary values of k

3 Switching Technologies

• Circuit Switching:

- also called connection-oriented network service
- each communication session consists of three phases:
 - establishment of a dedicated and exclusively used physical path between sender and receiver
 - data transfer
 - release of the connection
- all intermediate nodes work as simple switching elements
- all data elements use the same path from the sender to the receiver
- the total transmission time (for an m -byte message with a transmission rate of r bits/s is

$$T_{\text{total}} = T_{\text{transfer}} + T_{\text{transit}} + \frac{8m}{r} + T_{\text{release}}$$

- used for example in PSTN

3 Switching Technologies (cont'd)

• Packet Switching:

- also called connectionless network service
- each message is split into several packets (typically of a few thousand bits) that contain additional address and control information
- each packet may take a different route from the sender to the receiver
- consequently, the packets can have different latency and can arrive in a different order at the receiver node
- the packets are reassembled at the receiver to the original message
- if an m -byte message (split into p packets, each with a header of h bytes) is transmitted at a rate of r bits/s and the overhead for routing each packet in a node is T_{router} , then the total transmission time for a message over a network in k hops is

$$T_{\text{total}} = k \cdot T_{\text{router}} + (k + p) \cdot \left(T_{\text{transfer}} + \frac{8}{r} \left[\frac{m}{k} + h \right] \right)$$

- used for example in internet